

# New attacks on Keccak-224 and Keccak-256

Itai Dinur<sup>1</sup>, Orr Dunkelman<sup>1,2</sup> and Adi Shamir<sup>1</sup>

<sup>1</sup> Computer Science department, The Weizmann Institute, Rehovot, Israel

<sup>2</sup> Computer Science Department, University of Haifa, Israel

**Abstract.** The Keccak hash function is one of the five finalists in NIST’s SHA-3 competition, and so far it showed remarkable resistance against practical collision finding attacks: After several years of cryptanalysis and a lot of effort, the largest number of Keccak rounds for which actual collisions were found was only 2. In this paper we develop improved collision finding techniques which enable us to double this number. More precisely, we can now find within a few minutes on a single PC actual collisions in standard Keccak-224 and Keccak-256, where the only modification is to reduce their number of rounds to 4. When we apply our techniques to 5-round Keccak, we can get in a few days excellent near collisions, where the Hamming distance is 5 in the case of Keccak-224 and 10 in the case of Keccak-256. Our new attack combines differential and algebraic techniques, and uses the fact that each round of Keccak is only a quadratic mapping in order to efficiently find pairs of messages which follow a high probability differential characteristic.

**Keywords:** Cryptanalysis, SHA-3, Keccak, collision, near-collision, practical attack.

## 1 Introduction

The Keccak hash function [4] uses the sponge construction [3] to map arbitrary long inputs into fixed length outputs, and is one of the five finalists of NIST’s SHA-3 competition. The Keccak versions submitted to the SHA-3 competition have an internal state size of  $b = 1600$  bits, and an output size  $n$  of either 224, 256, 384 or 512 bits. The internal permutation of Keccak consists of 24 application of a non-linear round function, applied to the 1600-bit state. Previous papers on Keccak, such as [14], include analysis of Keccak versions with a reduced internal state size, or with different output sizes. However, in this paper, we concentrate on the standard Keccak versions submitted to the SHA-3 competition, and the only way in which we modify them is by reducing their number of rounds.

Previous results on Keccak’s internal permutation include zero-sum distinguishers presented in [1], and later improved in [5, 6, 9]. Although zero-sum distinguishers reach a significant number of rounds of Keccak’s internal permutation, they have very high complexities, and they seem unlikely to threaten the core security properties of Keccak (namely, collision resistance, preimage resistance and second-preimage resistance). Other results on Keccak’s internal permutation include a differential analysis given in [10]. Using techniques adapted

from the rebound attack [13], the authors construct differential characteristics which give distinguishers on up to 8 rounds of the permutation, with complexity of about  $2^{491}$ . However, in their method it is not clear how to reach the starting state differences of these characteristics from valid initial states of Keccak’s internal permutation, since in sponge constructions a large portion of the initial state of the permutation is fixed and cannot be chosen by the cryptanalyst. Thus, although the results of [10] seem to be more closely related to the core security properties of Keccak than zero-sum distinguishers, they still do not lead to any attacks on the Keccak hash function itself.

Currently, there are very few results that analyze reduced-round variants of the full Keccak (rather than its building blocks): in [2], Bernstein described preimage attacks which extend up to 8 rounds of Keccak, but are only marginally faster than exhaustive search, and use a huge amount of memory. More recently, Naya-Plasencia, Röck and Meier presented practical attacks on Keccak-224 and Keccak-256 with a very small number of rounds [15]. These attacks include a preimage attack on 2 rounds, as well as collisions on 2 rounds and near-collisions on 3 rounds. In this paper, we extend these collision attacks on Keccak-224 and Keccak-256 by 2 additional rounds: we find actual collisions in 4 rounds and actual near-collisions in 5 rounds of Keccak-224 and Keccak-256, with Hamming distance 5 and 10, respectively.

The collisions and near-collisions of [15] were obtained using low Hamming weight differential characteristics, starting from the initial state of Keccak’s permutation. Such low Hamming weight characteristics are also the starting point of our new attacks, but we do not require the characteristics to start from the initial state of the permutation. Given a low Hamming weight starting state difference of a characteristic, we can easily extend it backwards by one round, and maintain its high probability (as done in [10]). However, due to the very fast diffusion of the inverse linear mapping used by Keccak’s permutation, the new starting state difference of the extended characteristic has a very high Hamming weight. We call this starting state difference a *target difference*, since our goal is to find message pairs which have this difference after one round of the Keccak permutation (after the fixed round, this difference will evolve according to the characteristic with high probability).<sup>1</sup> One of the main tools we develop in this paper is an algorithm that aims to achieve this goal, namely, to find message pairs which satisfy a given target difference after one Keccak permutation round. We call this algorithm a *target difference algorithm*, and it allows us to extend our initial characteristic by two additional rounds (as shown in Figure 1): we first extend the characteristic backwards by one round to obtain the target difference (while maintaining the characteristic’s high probability). Then, we use the target difference algorithm to link the characteristic to the initial state of Keccak’s permutation, through an additional round. We note that the final link,

---

<sup>1</sup> We note that the target difference is not a valid initial difference of the permutation, which fixes many of the state bits to pre-defined values. As a result, the high probability characteristic cannot be used to extend the results of [15] by an additional round.

which efficiently bypasses Keccak’s first Sbox layer, uses algebraic techniques rather than standard probabilistic techniques.

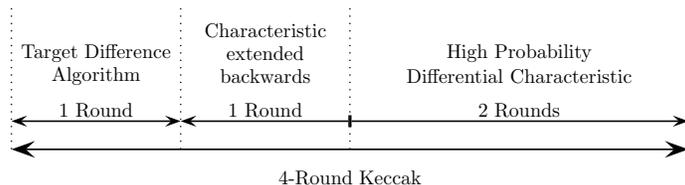
The target difference algorithm is related to several hash function cryptanalytic techniques that were developed in recent years. In particular, it is related to the work of Khovratovich, Biryukov and Nikolic [12], where, similarly to our algorithm, the authors use linear algebra to quickly satisfy many conditions of a differential characteristic. However, these techniques seem to work best on byte-oriented hash functions, whose internal structure can be described using a few sparse equations, which is not the case for Keccak. Our algorithm is also closely related to the work of Khovratovich [11] that exploits structures (which aggregate internal states of the hash function) in order to reduce the amortized complexity of collision attacks: the attacker first finds a truncated differential characteristic and searches for a few pairs of initial states that satisfy it. Then, using the structures and the initially found pairs, the attacker efficiently obtains many additional pairs that satisfy the truncated characteristic. However, in the case of Keccak, there are very few characteristics that can lead to a collision with high probability, and it seems unlikely that they can be joined in order to form the truncated differential characteristic required in order to organize the state differences into such structures. Moreover, it seems difficult to find even one pair of initial states that satisfy the target difference for Keccak. Another attack related to the target difference algorithm is the rebound attack [13]. In this attack, the cryptanalyst uses the available degrees of freedom to efficiently link and extend two truncated differential characteristics, both forwards and backwards, from an intermediate state of the hash function. However, once again, such high probability truncated characteristics are unlikely to exist for Keccak. Moreover, it is not clear how to use the rebound attack to link the backward characteristic to the initial state of the permutation. Thus, our target difference algorithm can be viewed as an asymmetric rebound attack, where one side of the characteristic is fixed.

Our full attacks have two parts, where in the first part we execute the target difference algorithm in order to obtain a sufficiently large set of message pairs that satisfy the target difference after the first round. In the second part of the attack, we try different message pairs in this set in order to find a pair whose difference evolves according to a characteristic whose starting state is the target difference. Since the target difference algorithm does not control the differences beyond the first round, the second part of the attack is a standard probabilistic differential attack (which only searches for collisions or near-collisions obtained from message pairs within a specific set). The high probability differential characteristic beyond the first round ensures that the time complexity of the second part of the attack is relatively low.

Although the target difference algorithm is heuristic, and there is no provable bound on its running time, it was successfully applied with its expected complexity to many target differences defined by the high probability differential characteristics. Consequently, we were able to find actual collisions for 4 rounds of Keccak-224 and Keccak-256 within minutes on a standard PC. By

using good differential characteristics for an additional round, we found near-collisions for 5 rounds of Keccak-224 and Keccak-256. However, this required more computational effort (namely, a few days on a single PC), since the extended characteristics have lower probabilities.

The paper is organized as follows. In Section 2, we briefly describe Keccak, and in Section 3 we introduce our notations. In Section 4, we give a comprehensive overview of the target difference algorithm and describe the properties of Keccak that it exploits. In Section 5, we present our results on round-reduced Keccak. In the full version of the paper [8], we describe the full details of the target difference algorithm, and propose an alternative algorithm, which has a better understood time complexity. Since the original algorithm gave us very good results in practice, we did not use this alternative version. However, it may be more efficient in some cases, especially if someone finds longer high probability characteristics for Keccak’s permutation.



**Fig. 1.** Extending a 2-Round Differential Characteristic by Two Additional Rounds

## 2 Description of Keccak

In this section we give short descriptions of the sponge construction and the Keccak hash function. More details can be found in the Keccak specification [4].

The sponge construction [3] works on a state of  $b$  bits, which is split into two parts: the first part contains the first  $r$  bits of the state (called the outer part of the state) and the second part contains the last  $c = b - r$  bits of the state (called the inner part of the state).

Given a message, it is first padded and cut into  $r$ -bit blocks, and the  $b$  state bits are initialized to zero. The sponge construction then processes the message in two phases: In the absorbing phase, the message blocks are processed iteratively by XORing each block into the first  $r$  bits of the current state, and then applying a fixed permutation on the value of the  $b$ -bit state. After processing all the blocks, the sponge construction switches to the squeezing phase. In this phase,  $n$  output bits are produced iteratively, where in each iteration the first  $r$  bits of the state are returned as output and the permutation is applied.

The Keccak hash function uses multi-rate padding: given a message, it first appends a single 1 bit. Then, it appends the minimum number of 0 bits followed

by a single 1 bit, such that the length of the result is a multiple of  $r$ . Thus, multi-rate padding appends at least 2 bits and at most  $r + 1$  bits.

The Keccak versions submitted to the SHA-3 competition have  $b = 1600$  and  $c = 2n$ , where  $n \in \{224, 256, 384, 512\}$ . The 1600-bit state can be viewed as a 3-dimensional array of bits,  $a[5][5][64]$ , and each state bit is associated with 3 integer coordinates,  $a[x][y][z]$ , where  $x$  and  $y$  are taken modulo 5, and  $z$  is taken modulo 64.

The Keccak permutation consists of 24 rounds, which operate on the 1600 state bits. Each round of the permutation consists of five mappings  $R = \iota \circ \chi \circ \pi \circ \rho \circ \theta$ . Keccak uses the following naming conventions, which are helpful in describing these mappings:

- A row is a set of 5 bits with constant  $y$  and  $z$  coordinates, i.e.  $a[*][y][z]$ .
- A column is a set of 5 bits with constant  $x$  and  $z$  coordinates, i.e.  $a[x][*][z]$ .
- A lane is a set of 64 bits with constant  $x$  and  $y$  coordinates, i.e.  $a[x][y][*]$ .
- A slice is a set of 25 bits with a constant  $z$  coordinate, i.e.  $a[*][*][z]$ .

The five mappings are given below, for each  $x, y$ , and  $z$  (where the state addition operations are over  $GF(2)$ ):

1.  $\theta$  is a linear map, which adds to each bit in a column, the parity of two other columns.

$$\theta: a[x][y][z] \leftarrow a[x][y][z] + \sum_{y'=0}^4 a[x-1][y'][z] + \sum_{y'=0}^4 a[x+1][y'][z-1]$$

In this paper, we also use the inverse mapping,  $\theta^{-1}$ , which is more complicated and provides much faster diffusion: for  $\theta^{-1}$ , flipping the value of any input bit, flips the value of more than half of the output bits.

2.  $\rho$  rotates the bits within each lane by  $T(x, y)$ , which is a predefined constant for each lane.

$$\rho: a[x][y][z] \leftarrow a[x][y][z + T(x, y)]$$

3.  $\pi$  reorders the lanes.

$$\pi: a[x][y][z] \leftarrow a[x'][y'][z], \text{ where } \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix} \cdot \begin{pmatrix} x' \\ y' \end{pmatrix}$$

4.  $\chi$  is the only non-linear mapping of Keccak, working on each of the 320 rows independently.

$$\chi: a[x][y][z] \leftarrow a[x][y][z] + ((\neg a[x+1][y][z]) \wedge a[x+2][y][z])$$

Since  $\chi$  works on each row independently, it can be viewed as an Sbox layer which simultaneously applies the same 5 bits to 5 bits Sbox to the 320 rows of the state. We note that the Sbox function is an invertible mapping, and we will use the extremely important observation that the algebraic degree of each output bit of  $\chi$  as a polynomial in the five input bits is only 2. As noted in [4], the algebraic degree of the inverse mapping  $\chi^{-1}$  is 3.

5.  $\iota$  adds a round constant to the state.

$$\iota: a \leftarrow a + RC[i_r]$$

We omit the values of  $RC[i_r]$ , as they are not needed for our analysis.

### 3 Notations

Given a message  $M$ , we denote its length in bits by  $|M|$ . Unless specified otherwise, in this paper we assume that  $|M| = r - 8$ , namely we consider only single-block messages of maximal length such that  $|M| \pmod{8} \equiv 0$  (which give us the maximal number of degrees of freedom, for single-block messages containing an integral number of bytes). Given  $\overline{M}$ , we denote the initial state of the Keccak permutation as the 1600-bit word  $\overline{M} \triangleq M || p || 0^{2n}$ , where  $||$  denotes concatenation, and  $p$  denotes the 8-bit pad 10000001.

The first three operations of Keccak's round function are linear mappings, and we denote their composition by  $L \triangleq \rho \circ \pi \circ \theta$ . We denote the Keccak nonlinear function on 5-bit words defined by varying the first index by  $\chi_{|5}$ . The difference distribution table (*DDT*) of this function is a two-dimensional  $32 \times 32$  integer table, where all the differences are assumed to be over  $GF(2)$ . The entry  $DDT(\delta^{in}, \delta^{out})$  specifies the number of input pairs to this Sbox with difference  $\delta^{in}$  that give the output difference  $\delta^{out}$  (i.e., the size of the set  $\{x \in \{0, 1\}^5 \mid \chi_{|5}(x) + \chi_{|5}(x + \delta^{in}) = \delta^{out}\}$ ).

We denote the 1600-bit target difference, which is the input of the target difference algorithm, by  $\Delta_T$ . The output of the algorithm is a subset of ordered pairs of single block messages  $\{(M_1^1, M_1^2), (M_2^1, M_2^2), \dots, (M_k^1, M_k^2)\}$  that satisfy this difference after one round  $R$ , namely  $R(\overline{M}_i^1) + R(\overline{M}_i^2) = \Delta_T \forall i \in \{1, 2, \dots, k\}$ .

## 4 Overview of the Target Difference Algorithm

When designing the target difference algorithm, we face two problems: first, the target difference extends backwards, beyond the first Keccak Sbox layer, with very low probability (due to its high Hamming weight). The second problem is that the initial state of the permutation fixes many of the state bits to pre-defined values, and the initial states that we use must satisfy these constraints. On the other hand, Keccak has several useful properties that we can exploit in our target difference algorithm. In this section, we describe these properties in detail and give an overview of the algorithm.

### 4.1 The Properties of Keccak Exploited by the Target Difference Algorithm

**The First Property** Keccak-224 and Keccak-256 allow the user to control many of the 1600 state bits of the initial state of the permutation. Thus, given a target difference, we expect many solutions to exist (namely, one-block message pairs which have the 1600-bit target difference after one permutation round): since we consider message pairs, where each message is of length  $r - 8 = 1600 - 8 - 2n$  bits (1144 for Keccak-224, and 1080 for Keccak-256), given an arbitrary 1600-bit target difference, there is an expected number of  $2^{2^{(1600-8-2n)-1600}} = 2^{1584-4n}$  message pairs of this length that satisfy this difference (regardless of

the value of the inner part of the state). Thus, the algorithm has 704 and 560 degrees of freedom for Keccak-224 and Keccak-256, respectively.

Despite the large number of available degrees of freedom, the number of possible solutions varies significantly according to the target difference. To demonstrate this, we use the fact that  $L^{-1}$  has very fast diffusion (i.e., even an input with one non-zero bit is mapped by  $L^{-1}$  into a roughly balanced output). We consider the case where  $t > 0$  out of the 320 Sboxes of the target difference are active (i.e., they have a non-zero output difference). Each one of the  $320 - t$  non-active Sbox zero output differences is uniquely mapped backwards to a zero input difference into the first Sbox layer. Using the Keccak Sbox  $DDT$ , it is easy to see that each one of the  $t$  active Sbox output differences is mapped to more than 8 possible input differences. Thus, the number of possible state differences after the first linear layer (or before the first Sbox layer) is more than  $8^t = 2^{3t}$ . Since  $L$  is invertible and acts deterministically on the differences, the number of possible input differences to the Keccak compression function remains the same. We now recall from the first difference constraint in Section 4, that we require that the  $2n + 8$  MSBs of  $\Delta_I$  are zero. However, for  $t$  large enough, we still expect more than  $2^{3t-2n-8}$  valid solutions. When the target difference is chosen at random, we have  $t \approx 310$  (since the probability that an Sbox output difference is zero is  $\frac{1}{32}$ ). This gives more than  $2^{930-448-8} = 2^{474}$  expected solutions for Keccak-224, and more than  $2^{930-512-8} = 2^{410}$  expected solutions for Keccak-256. On the other hand, consider the extreme case of  $t = 1$  (i.e., the target difference has only one active Sbox). Clearly, this Sbox cannot contribute more than 31 possible differences after the first linear layer. Since  $L^{-1}$  has very fast diffusion, these possible differences are mapped to at most 31 roughly balanced non-zero possible input differences, and we do not expect the  $2n+8$  MSBs of any of them to be zero. To conclude, target differences with a small number of active Sboxes are likely to have no solutions at all. On the other hand, a majority of the target differences have a very large number of expected solutions for Keccak-224 and Keccak-256. Note that having a large number of solutions does not imply that it is easy to find any one of them, since their density is still minuscule.

**The Second Property** The algebraic degree of the Keccak Sboxes is only 2. This implies that given a 5-bit input difference  $\delta^{in}$  and a 5-bit output difference  $\delta^{out}$ , the set of values  $\{v_1, v_2, \dots, v_l\}$  such that  $\chi_{|5}(v_i) + \chi_{|5}(v_i + \delta^{in}) = \delta^{out}$  is an affine subset. Since  $(v_i + \delta^{in}) + \delta^{in} = v_i$ , then  $v_i + \delta^{in} \in \{v_1, v_2, \dots, v_l\}$ , implying  $\{v_1, v_2, \dots, v_l\} = \{v_1 + \delta^{in}, v_2 + \delta^{in}, \dots, v_l + \delta^{in}\}$ . Thus, both coordinates of the ordered pairs give the same subset, and we denote it by  $A(\delta^{in}, \delta^{out})$  (note that  $|A(\delta^{in}, \delta^{out})| = DDT(\delta^{in}, \delta^{out})$ ). On the other hand, since the algebraic degree of the inverse Sbox is 3, which is reduced to 2 (rather than 1) after differentiation, the output values that satisfy an input and an output difference do not necessarily form an affine subset.

**The Third Property** For any non-zero 5-bit output difference  $\delta^{out}$  to a Keccak Sbox, the set of possible input differences,  $\{\delta^{in} | DDT(\delta^{in}, \delta^{out}) > 0\}$ , contains at

least 5 (and up to 17) 2-dimensional affine subspaces. These affine subspaces can be easily pre-computed using the *DDT*, for each one of the 31 possible non-zero output differences. However, we note that there is no output difference for which the set of possible input differences contains an affine subspace of dimension 3 or higher.

## 4.2 Formulating the Problem

Given  $\Delta_T$ , an arbitrary message pair  $(M^1, M^2)$  in which  $|M^1| = |M^2| = r - 8$  is a solution to our problem if  $R(\overline{M}^1) + R(\overline{M}^2) = \Delta_T$ . This can be formulated using two constraints on the 1600-bit words  $(\overline{M}_1, \overline{M}_2)$ :

1. The  $2n + 8$  MSBs of  $\overline{M}^1$  and  $\overline{M}^2$  are equal to  $p||0^{2n}$ , where  $p$  denotes the 8-bit pad 10000001.
2.  $R(\overline{M}^1) + R(\overline{M}^2) = \Delta_T$  (where  $R$  is the permutation round of Keccak).

We can easily formulate the first constraint using linear equations on the bits of  $\overline{M}_1$  and  $\overline{M}_2$ . Since Keccak’s Sbox has an algebraic degree of 2 over  $GF(2)$ , we can formulate the second constraint as a system of quadratic equations on these bits. Standard heuristic techniques for solving such systems include using the available degrees of freedom to fix some message values (or values before the first Sbox layer) in order to linearize the system. However, these techniques require many more than the available number of degrees of freedom. For example, in order to get linear equations after one round of Keccak’s permutation, we can fix 3 out of the 5 bits entering an Sbox (after the first linear layer), such that there are no two consecutive unknown input bits entering the Sbox. Using this technique reduces the single quadratic term in the symbolic form of each of the Sbox’s output bits to a linear term. However, this requires fixing  $320 \cdot 3 = 960$  bits per message, and  $2 \cdot 960 = 1920$  bits in total, which is significantly more than the 704 available degrees of freedom for Keccak-224 (and clearly more than the available number of degrees of freedom for the other Keccak versions). Consequently, we have to repeat the linearization procedure a huge number of times, with different fixed values, in order to find a solution.

**A Two-Phase Algorithm** Although we expect our quadratic system to have many solutions, solving all the equations at once seems difficult. Thus, we split the problem into easier tasks by exploiting the low algebraic degree of Keccak’s Sbox to a greater extent than in the standard techniques: as described in the second property of Section 4.1 given an input difference and an output difference to an Sbox, all the pairs of input values that satisfy them form an affine subset.<sup>2</sup> This suggests an algorithm with two phases, where in the first phase (called the

<sup>2</sup> Similar observations were used in [7] to suggest that when  $DDT(\delta^{in}, \delta^{out}) = 2$  or 4, the same holds. In the specific case of Keccak, we also use 3-dimensional affine subsets of pairs that satisfy the Sbox difference transition  $(\delta^{in}, \delta^{out})$ , for which  $DDT(\delta^{in}, \delta^{out}) = 8$ .

*difference phase*) we find an input difference to all the Sboxes, and in the second phase (called the *value phase*) we obtain the actual values of the message pairs that lead to the target difference.

Using this two-phase approach, the ordered pairs produced by our algorithm satisfy two additional properties: the 1600-bit *input difference* of the initial states is fixed to some 1600-bit value  $\Delta_I$  (i.e.  $\overline{M}_i^1 + \overline{M}_i^2 = \Delta_I \forall i \in \{1, 2, \dots, k\}$ ), and the set composed of all the initial states defined by the first message in each ordered pair (i.e.  $\bigcup\{\overline{M}_i^1\} \forall i \in \{1, 2, \dots, k\}$ ), forms an affine subset. The algorithm outputs the ordered pairs as the fixed 1600-bit input difference  $\Delta_I$ , and some basis for the affine subset  $\bigcup\{\overline{M}_i^1\} \forall i \in \{1, 2, \dots, k\}$ . We note that the large number of degrees of freedom allows us to restrict the set of solutions (i.e. the set of message pairs that satisfy the target difference) to a smaller subset (but still large enough for our purposes) that can be found relatively easily. In particular, the algorithm considers only message pairs with a fixed difference  $\Delta_I$ , for which all the solutions can be found by solving linear equations.

The two constraints above, which define our quadratic equation system, are broken into two sets of constraints, since we have to simultaneously enforce two *difference constraints* (given as constraints on the 1600-bit word  $\Delta_I$ ):

**Difference Constraint 1** *The  $2n + 8$  most significant bits (MSBs) of  $\Delta_I$  are equal to zero.*

**Difference Constraint 2**  *$L(\Delta_I)$  is a valid input difference to the Sbox layer, i.e. there exists some 1600-bit word  $W$  such that  $\chi(W) + \chi(W + L(\Delta_I)) = \Delta_T$  (note that since  $L$  is a linear function,  $L(\Delta_I)$  is well-defined).*

The first difference constraint simply equates bits of the input difference  $\Delta_I$  to zero (456 bits for Keccak-224 and 520 bits for Keccak-256), while the second difference constraint assigns to every 5 bits of  $L(\Delta_I)$  that enter an Sbox, several possible values which are not related by simple affine equations.

In the second phase, we enforce additional *value constraints* (given on the 1600-bit word  $\overline{M}^1$ ):

**Value Constraint 1** *The  $2n + 8$  MSBs of  $\overline{M}^1$  are equal to  $p||0^{2n}$ , where  $p$  denotes the 8-bit pad 10000001.*

**Value Constraint 2**  *$R(\overline{M}^1) + R(\overline{M}^1 + \Delta_I) = \Delta_T$ .*

Note that the first difference constraint and the first value constraint on each  $\overline{M}_i^1$  also ensure that the same value constraint holds for  $\overline{M}_i^2$  (i.e., the  $2n + 8$  MSBs of  $\overline{M}_i^2$  are equal to  $p||0^{2n}$ ).

Given a single 1600-bit Sbox layer input difference, the second property of Section 4.1 implied that enforcing the two value constraints simply reduces to solving a union of two sets of linear equations. On the other hand, it is not clear how to simultaneously enforce both of the difference constraints, since given an

output difference to an Sbox  $\delta^{out}$ , all the possible input differences  $\delta^{in}$  such that  $DDT(\delta^{in}, \delta^{out}) > 0$ , are not related by simple affine relations.

### 4.3 The Difference Phase

**Unsuccessful Attempts to Enforce the Difference Constraints** We can try to enforce both difference constraints by assigning the undetermined  $1600 - 2n - 8$  bits of  $\Delta_I$ , in such a way that the second difference constraint will hold. This usually involves iteratively constructing an assignment for  $\Delta_I$ , by guessing several undetermined bits at a time, and filtering the guesses by verifying the second difference constraint. However, this is likely to have a very large time complexity, since  $L$  diffuses the bits of  $\Delta_I$  in a way that forces us to guess many bits before we can start filtering the guesses. Moreover, for any  $\Delta_I$ , the fraction of input differences satisfying the first difference constraint that also satisfy the second difference constraint is very small. Thus, most of the computational effort turns out to be useless, since the guesses are likely to be discarded at later stages of the algorithm. Another approach is to guess  $L(\Delta_I)$  by iteratively guessing the 5-bit Sbox input differences, and filtering the guesses by verifying the first difference constraint. For similar reasons, this approach is likely to have a very large time complexity.

**A Better Approach** Both of these approaches are very strict, since each guess made by the algorithm commits to a specific value for some of the bits of  $\Delta_I$ , or  $L(\Delta_I)$ , and restricts the solution space significantly. Thus, we use the third property of Section 4.1, which gives us more flexibility, and significantly reduces the time complexity: given any non-zero 5-bit output difference to a Keccak Sbox, the set of possible input differences contains at least five 2-dimensional affine subspaces. Consequently, in order to enforce the second difference constraint, for each Sbox with a non-zero output difference (i.e., an active Sbox), we choose one of the affine subsets (which contains 4 potential values for the 5 Sbox input bits of  $L(\Delta_I)$ ), instead of choosing specific values for these bits. This enables us to maintain an affine subspace of potential values for  $L(\Delta_I)$ , starting with the full 1600-dimensional space, and iteratively reducing its dimension by adding affine equations in order to enforce the second difference constraint for each Sbox. In addition to these affine equations that we add per active Sbox, we also have to add the linear equations for the non-active Sboxes (which equate their 5 input difference bits to zero), and the additional  $2n + 8$  linear equations that enforce the first difference constraint. All of these equations are added to a linear system of equations that we denote by  $E_\Delta$ .

Since the  $2n + 8$  equations that enforce the first difference constraint do not depend on the target difference, we add them to  $E_\Delta$  before we iterate the Sboxes. While iterating over the active Sboxes, we add equations on  $L(\Delta_I)$  in order to enforce the second difference constraint and hope that for each Sbox, we can add equations such that  $E_\Delta$  is consistent. Note that the equations in  $E_\Delta$  in each stage of the algorithm depend on the order in which we consider the

active Sboxes, and on the order in which we consider the possible affine subsets of input differences for each Sbox. Thus, if we reach an Sbox for which we cannot add equations in order to enforce the second constraint (while maintaining the consistency of  $E_\Delta$ ), it does not imply that it is impossible to satisfy the difference constraints. In this case, we can simply change the order in which we consider the active Sboxes, or the order in which we consider the affine subsets for each Sbox, and try again. Since we cannot predict in advance the orderings that give the best result, we choose them heuristically, as described in the full version of the paper [8].

#### 4.4 The Value Phase

In case the difference phase procedure described above succeeds, it actually outputs an affine subspace of candidate input differences, rather than a single value for  $\Delta_I$ . Next, we can commit to a specific value for  $\Delta_I$  and run the value phase, hoping that the set of all linear equations defined by the value constraints has a solution. Namely, we allocate another system of equations, which we denote by  $E_M$ , and add the equations on  $\overline{M}^1$  that enforce the first value constraint. We then add the additional linear equations that enforce the second value constraints for all the Sboxes, and output the solution to the system, if it exists. However, once again, this approach is too strict, and may force us to repeat the value phase a huge number of times with different values for  $\Delta_I$ , until we find a solution. Thus, we do not choose a single value for  $\Delta_I$  in advance. Instead, we reduce the linear subset of candidates for  $\Delta_I$  gradually by fixing the input difference to each one of the active Sboxes, until a single value for  $\Delta_I$  remains. Thus, we continue to maintain  $E_\Delta$  throughout the value phase, and iteratively add the additional 2 equations which are required to uniquely specify a 5-bit input difference for each active Sbox, among the 2-dimensional affine subsets chosen in the difference phase. Once we fix the input difference to an Sbox, we immediately obtain linear equations on  $\overline{M}^1$ , and we can check their consistency with the current equations in  $E_M$ . In case the equations in  $E_M$  are not consistent for a certain Sbox, we can try to choose another input difference for it. This gives different equations on  $\overline{M}^1$ , which may be consistent and allow us to continue the process.

Similarly to the difference phase, the equations in  $E_M$  in each stage of the algorithm depend on the order in which we consider the active Sboxes, and on the order in which we consider the possible input differences for each Sbox. Thus, once again, if at some stage of the value phase we cannot add any consistent equations to  $E_M$ , we can change one of these orderings and try again, hoping to obtain a valid solution.

We stress again that both phases of the algorithm are not guaranteed to succeed. The success of each phase depends on the target difference, and on orderings which are chosen heuristically. As a result, we may have to iterate both phases of the algorithm an undetermined number of times with modified orderings, hoping to obtain better results.

## 5 Application of the Target Difference Algorithm to Round-Reduced Keccak

Since we would like to use the target difference algorithm in order to find collisions and near-collisions in Keccak, it is crucial to verify the algorithm’s success on target differences which lead to these results. Thus, before we run the algorithm, we have to find such high probability differential characteristics, and to obtain the target differences which are likely to be the most successful inputs to the algorithm. As described in the introduction, once we find a high probability differential characteristic with a low Hamming weight starting state difference, we extend it backwards to obtain the target difference (while maintaining its high probability). We then use the target difference algorithm to link the extended characteristic backwards to the initial state of Keccak’s permutation, with an additional round. Thus, any low Hamming weight characteristic for  $r$  rounds of Keccak’s permutation can be used to obtain results on a round-reduced version of Keccak with  $r + 2$  round. Specifically, in this section we demonstrate how we use 2-round characteristics in order to find collisions for 4 rounds of Keccak-224 and Keccak-256, and how to use 3-round characteristics in order to find near-collisions for 5 rounds of these Keccak versions.

### 5.1 Searching for Differential Characteristics

We reuse the notion of a *column parity kernel* or *CP-kernel* that was defined in the Keccak submission document [4]: a 1600-bit state difference is in the CP-kernel if all of its columns have even parity. It is easy to see that such state differences are fixed points of the function  $\theta$ , which does not increase their Hamming weight. Since  $\rho$  and  $\pi$  just reorder the bits of the state, the application of  $L$  to a CP-kernel does not change its total Hamming weight. In addition, there is a high probability that such low Hamming weight differential states are fixed points of  $\chi$ . Thus, when we start a differential characteristic from a low Hamming weight CP-kernel, we can extend it beyond the Sbox layer,  $\chi$ , to one additional round of the Keccak permutation, with relatively high probability and without increasing its Hamming weight. However, extending such a characteristic to more rounds in a similar way is more challenging, since we have to ensure that the state difference before the application of  $\theta$  remains in the CP-kernel at the beginning of each round.

**Using Previous Results** In [10] and [15], the authors propose algorithms for constructing low Hamming weight differential characteristics for Keccak. Both of these algorithms successfully find differential characteristics that stay in the CP-kernel for 2 rounds (named *double kernel trails* in [15]), some of which lead to collisions on the  $n$ -bit extract taken from the final state after 2 rounds, with high probability. However, when trying to extend each one of these characteristics by another round, the state difference is no longer in the CP-kernel and thus its Hamming weight increases significantly (from less than 10 to a few dozen

bits). Nevertheless, the Hamming weight of the characteristics is still relatively low, and they can lead with reasonably high probability to near-collisions on the  $n$  output bits extracted. Beyond 3 rounds, the Hamming weight of the characteristics becomes very high (more than 100), and it seems unlikely that they can be extended to give collisions or near-collisions with reasonable probability. The currently known double kernel differential trails only extend forward to at most three rounds with reasonably high probability (higher than  $2^{-100}$ ). Finding new high probability differential characteristics, starting from a low Hamming weight state difference and extending forwards more than 3 rounds, remains a challenging task, which we do not deal with in this paper.

Our attacks on round-reduced Keccak make use of the type of differential characteristics that were found in [10] and [15], namely low Hamming weight characteristics that stay in the CP-kernel for 2 rounds. The double kernel trails with the highest probability have Hamming weight of 6 at the input to the initial round, and due to their low hamming weight, we could easily find all these characteristics within a minute on a standard PC. There are 571 such characteristics out of which, 128 can give collisions for Keccak-224 and 64 can give collisions for Keccak-256. However, when trying to extend the characteristics by an additional round, we were not able to find any characteristic that gives collisions for Keccak-224 (or Keccak-256) with reasonable probability. Thus, our best 3-round characteristics lead only to near-collisions, rather than collisions. The characteristics that give the near-collisions with the smallest difference Hamming weight for Keccak-224 and Keccak-256 are, again, double kernel trails with 6 non-zero input bits. The best 3-round characteristics for Keccak-224 lead to near-collisions with a difference Hamming weight of 5, and for Keccak-256, the best 3-round characteristics leads to a near-collision difference Hamming weight of 8. Examples of these characteristics are found in Appendix A.

**Extending the characteristics Backwards** Since the characteristics that we use start with a low Hamming weight state difference, we can extend them backwards by one round without reducing their probability significantly (as done in [10]): we take this low Hamming weight initial state difference, and choose a valid state difference input to the previous Sbox layer which could produce it. We then apply  $L^{-1}$ , and obtain a new initial state difference for the extended characteristic, which serves as a target difference for our new algorithm. Note that the target difference is not in the CP-kernel (otherwise, we would have found a low Hamming weight differential characteristic that stays in the CP-kernel for 3 rounds). Thus, when we apply  $L^{-1}$  to the state difference that enters the Sbox layer, we usually obtain a roughly balanced target difference, with only a few non-active Sboxes. This is significant to the success of the target difference algorithm, which strongly depends on the number of active Sboxes in the target difference.<sup>3</sup> In case the target difference obtained from a characteristic has too

---

<sup>3</sup> As demonstrated in Section 4.1, we expect a large number of non-active Sboxes to foil the target difference algorithm. This should be contrasted to differential attacks,

many non-active Sboxes, we can try to select another target difference for the characteristic, by tweaking the state difference input to the second Sbox layer.

Assuming that the algorithm succeeds and we obtain a sufficiently large linear subspace of message pairs (such that it contains at least one pair whose difference evolve according to the characteristic), we can find collisions for 4 rounds and near-collisions for 5 rounds of Keccak-224 and Keccak-256. For example, given an extended characteristic which results in collisions for 3 round of Keccak-256 with probability  $2^{-24}$ , we need a linear subspace containing at least  $2^{24}$  message pairs in order to find a collision on 4-round Keccak-256 with high probability.

## 5.2 Applying the Target Difference Algorithm to the Selected Differential Characteristics

We tested our target difference algorithm using a standard PC, on dozens of double-kernel trails with Hamming weight of 6. For each one of them, after tweaking the state difference input to the second Sbox layer at most once, we could easily compute a target difference where all of the 320 Sboxes are active. We then ran the target difference algorithm on each one of these targets. For both Keccak-224 and Keccak-256, the target difference algorithm eventually succeeded: the basic procedure of the difference phase always succeeded within the first two attempts (after changing the order in which we considered the Sboxes), while the value phase was more problematic, and we had to iterate its basic procedure dozens to thousands of times in order to find a good ordering of the Sboxes and obtain results. For Keccak-224, the algorithm typically returned an affine subspace of message pairs with a dimension of about 100 within one minute. For Keccak-256, the dimension of the affine subspaces of message pairs returned was typically between 35 and 50, which is smaller compared to the typical result size for Keccak-224 (as expected since we have fewer degrees of freedom). In addition, unlike Keccak-224, for Keccak-256 we had to rerun the algorithm (starting from the difference phase) a few times, when the value phase did not seem to succeed for the choice of candidate input difference subset. Hence, the running time of the algorithm was typically longer – between 3 and 5 minutes, which is completely practical.

## 5.3 Obtaining Actual Collisions and Near-Collisions for Round-Reduced Keccak-224 and Keccak-256

**Obtaining Collisions** After successfully running the target difference algorithm, we were able to find collisions for 4-round Keccak for each tested double-kernel trail with Hamming weight of 6 (which leads to a collision). Since the probability of each one of these differential characteristics is greater than  $2^{-30}$ , the probability that a random pair which satisfies its corresponding target difference leads to a collision, is greater than  $2^{-30}$ . Thus, we expect to find collisions

---

where the attacker searches for differential characteristics with many non-active Sboxes, which ensure that the differential transitions occur with high probability.

quickly for both Keccak-224 and Keccak-256, once the target difference algorithm returns a set of more than  $2^{30}$  message pairs. However, even though the subsets we used contained more than  $2^{30}$  message pairs, we were not able to find collisions within several of these subsets for Keccak-224, and for many of the subsets for Keccak-256. As a result, we had to rerun the target difference algorithm and obtain additional sets of message pairs, until a collision was found. Thus, the entire process of finding a collision typically takes about 2–3 minutes for Keccak-224, and 15–30 minutes for Keccak-256. The reason that there were no 4-round collisions within many of the message pair subsets, is the incomplete diffusion of the Keccak permutation within the first two rounds. Since our subsets of message pairs are relatively small (especially for Keccak-256), and the values of all the message pairs within a subset are closely related, some close relations between a small number of bits still hold before the Sbox layer of the second round (e.g., the value of a certain bit is always 0, or the XOR of two bits is always 1). Some of these relations make the desired difference transition into the second Sbox layer impossible, for all the message pairs within a subset. We note that we were still able to find collisions rather quickly, since it is easy to detect the cases where the difference transition within the second Sbox layer is impossible<sup>4</sup> (which allowed us to immediately rerun the target difference algorithm). In addition, when this difference transition is possible, we were always able to find collisions within the subset. Two concrete examples of colliding message pairs for Keccak-224 and Keccak-256 are given in Appendix B.

**Obtaining Near-Collisions** In order to obtain near-collisions on 5-round Keccak-224 and Keccak-256, we again start by choosing suitable differential characteristics. Out of all the characteristics that we searched, we chose the differential characteristics described in Appendix A, which lead to near-collisions of minimal Hamming weight for the two versions of Keccak. The results of the target difference algorithm when applied to targets chosen according to these characteristics, were similar to the results described in Section 5.2. However, compared to the probability of the characteristics leading to a collision, the probability of these longer characteristics is lower: the probability of the characteristics are  $2^{-57}$  and  $2^{-59}$  for Keccak-224 and Keccak-256, respectively. Thus, obtaining message pairs whose differences propagate according to these characteristics, and lead to 5-round near-collisions, is more difficult than obtaining collisions for 4 rounds of Keccak-224 and Keccak-256. However, for each such main characteristic, there are several secondary characteristics which diverge from the main one in final two rounds and give similar results. Thus, the probabilities of finding near collisions with a small Hamming distance for 5 rounds of Keccak-224

---

<sup>4</sup> In order to detect that the difference transition within the second Sbox layer is impossible for all the pairs in our subset, we try several arbitrary pairs in the subset, and observe if at least one has the desired difference after two rounds. Since we only need to check one Sbox layer transition, we expect that if this transition is indeed possible, we will find a corresponding message pair very quickly. Otherwise, we have to find a different set of message pairs by running the difference phase again.

and Keccak-256, are higher than the ones stated above. In addition, by using some simple message modification techniques within the subsets returned by the target difference algorithm, we were able to improve these probabilities further. Thus, for Keccak-224, we obtained near-collisions with a Hamming distance of 5, which is the same as the output Hamming distance of the main characteristic that we used. For Keccak-256, the main characteristic that we used has an output Hamming distance of 8, but we were only able to find message pairs which give a near-collision with a slightly higher Hamming distance of 10. All of these near-collisions were found within a few days on a standard PC. Examples of such near-collisions are given in Appendix B.

## 6 Conclusions and Future Work

In this paper, we presented practical collision and near-collision attacks on reduced-round variants of Keccak-224 and Keccak-256. Our attacks are based on a novel target difference algorithm, which is used to link high probability differential characteristics for the Keccak internal permutation to legal initial states of the hash function. Consequently, we were able to significantly improve the best known previous results on Keccak, by doubling (from 2 to 4) the number of rounds for which collisions can be found in a practical amount of time.

Our target difference algorithm is clearly limited by the number of available degrees of freedom, and it seems difficult to extend it to reach target differences spanning 2 or more rounds of the Keccak permutation. However, it seems very likely that the algorithm will be useful in the future if longer high probability differential characteristics are found for the Keccak permutation.

## References

1. Jean-Philippe Aumasson and Willi Meier. Zero-sum distinguishers for reduced Keccak-f and for the core functions of Luffa and Hamsi. NIST mailing list, 2009.
2. Daniel J. Bernstein. Second preimages for 6 (7? (8??)) rounds of keccak? NIST mailing list, 2010.
3. Guido Bertoni, Joan Daemen, Michaël Peeters, and Gilles Van Assche. Sponge functions. Presented at the ECRYPT Hash Workshop, 2007.
4. Guido Bertoni, Joan Daemen, Michaël Peeters, and Gilles Van Assche. The Keccak SHA-3 submission. Submission to NIST (Round 3), 2011.
5. Christina Boura and Anne Canteaut. Zero-Sum Distinguishers for Iterated Permutations and Application to Keccak-f and Hamsi-256. In Alex Biryukov, Guang Gong, and Douglas R. Stinson, editors, *Selected Areas in Cryptography*, volume 6544 of *Lecture Notes in Computer Science*, pages 1–17. Springer, 2010.
6. Christina Boura, Anne Canteaut, and Christophe De Cannière. Higher-order differential properties of keccak and luffa. Cryptology ePrint Archive, Report 2010/589, 2010. <http://eprint.iacr.org/>.
7. Joan Daemen and Vincent Rijmen. Plateau Characteristics. *IET Information Security*, 1(1):11–17, 2007.

8. Itai Dinur, Orr Dunkelman, and Adi Shamir. New attacks on Keccak-224 and Keccak-256. Cryptology ePrint Archive, Report 2011/624, 2011. <http://eprint.iacr.org/>.
9. Ming Duan and Xujia Lai. Improved zero-sum distinguisher for full round Keccak-f permutation. Cryptology ePrint Archive, Report 2011/023, 2011.
10. Alexandre Duc, Jian Guo, Thomas Peyrin, and Lei Wei. Unaligned rebound attack - application to keccak. Cryptology ePrint Archive, Report 2011/420, 2011.
11. Dmitry Khovratovich. Cryptanalysis of Hash Functions with Structures. In Michael J. Jacobson Jr., Vincent Rijmen, and Reihaneh Safavi-Naini, editors, *Selected Areas in Cryptography*, volume 5867 of *Lecture Notes in Computer Science*, pages 108–125. Springer, 2009.
12. Dmitry Khovratovich, Alex Biryukov, and Ivica Nikolic. Speeding up Collision Search for Byte-Oriented Hash Functions. In Marc Fischlin, editor, *CT-RSA*, volume 5473 of *Lecture Notes in Computer Science*, pages 164–181. Springer, 2009.
13. Florian Mendel, Christian Rechberger, Martin Schl affer, and S oren S. Thomsen. The Rebound Attack: Cryptanalysis of Reduced Whirlpool and Gr ostl. In Orr Dunkelman, editor, *FSE*, volume 5665 of *Lecture Notes in Computer Science*, pages 260–276. Springer, 2009.
14. Pawel Morawiecki and Marian Srebrny. A SAT-based preimage analysis of reduced KECCAK hash functions. Cryptology ePrint Archive, Report 2010/285, 2010.
15. Mar a Naya-Plasencia, Andrea R ock, and Willi Meier. Practical Analysis of Reduced-Round Keccak. In Daniel J. Bernstein and Sanjit Chatterjee, editors, *Progress in Cryptology - INDOCRYPT 2011*, volume 7107 of *LNCS*. Springer, Heidelberg, 2011.

## A Appendix: Differential Characteristics for Keccak

In this section, we give examples of 3-round differential characteristics, which lead to collisions on 4-round Keccak-224 and Keccak-256, and 4-round characteristics, which lead to near-collisions on 5-round Keccak-224 and Keccak-256.

The differential characteristics are given as a sequence of the starting state differences in each round. In all the presented characteristics, all the active Sboxes get an input difference with a Hamming weight of 1, and we assume that they produce the same differences as outputs (which occurs with probability  $2^{-2}$ ). In order to calculate the probability of the final transition, we only consider active Sboxes which effect the output bits (since we truncate the final state to obtain the hashed output). Each state difference is given as a matrix of  $5 \times 5$  lanes of 64 bits, ordered from left to right, where each lane is given in hexadecimal using the little-endian format. The symbol ‘-’ is used in order to denote a zero 4-bit difference value. For example, consider the second state difference in Characteristic 1: each of the first two lanes has a zero difference, and only the LSB of the third lane contains a non-zero difference.

## B Appendix: Actual Collisions and Near-Collisions for Round-Reduced Keccak-224 and Keccak-256

We give several examples of collisions and near-collisions for Keccak-224 and Keccak-256. The padded messages and output values are given in blocks of 32-

```
|26978AF134CB835E|AF224C4D78366789|C4DAE35E2656F26B|357C4789AF3-6AF1|78D3526BC6A74C4D|
|26978AF134CB835E|AF224C4D78366789|C4DAE35E2656F26B|357C4789AF3-6AF1|78D3526BC6A74C4D|
|26978AF134CB835E|AF224C4D78366789|C4DAE35E2676F26B|357C4789AF3-6AF1|78D3526BC4A74C4D|
|26978AF134CB835E|AF224C4D78366789|C4DAE35E2656F26B|357C4789AF3-4AF1|78D3526BC6A74C4D|
|26978AF134CB835E|AF226C4D78366789|C4DAE35E2656F26B|35FC4789AF3-6AF1|78D3526BC6A74C4D|
```

```
|-----|-----|-----1|-----4-----|-----|
|-----|-----|-----|-----|-----|
|-----|-----|-----|-----4-----|-----8-----|
|-----|-----|-----1|-----|-----8-----|
```

```
|-----|-----|-----|--8-----|2-----|
|4-----|-----|-----|-----|2-----|
|-----|-----|-----|--8-----|-----|
|4-----|-----|-----|-----|-----|
```

```
|-----|-----|-----|-----|-----|
|-----8-----|-----2-----|-----|-----|-----|
|-----|-----|-----1-----|-----|-----1-----|
|-----1-----|-----4-----|-----|-----|-----|
```

The probability of each one of the first two transitions is  $2^{-12}$ . The probability of the third transition is 1, since there are no active Sboxes which affect the output.

**Characteristic 1:** A 3-round characteristic leading to collisions on Keccak-224 and Keccak-256 with probability  $2^{-24}$

```
|BD135E2FA6BD1346|12D789A92F12D78F|D7E26BC344D7E224|E69AF134B5E69AD5|98BC4D6BF898BC58|
|BD135E2FA6BD1346|12D789A82F12D78F|D7E26BC344D7E264|E69AF134B5E69AD5|98BC4D6BF898BC58|
|BD135E2FA6BD1346|12D789AB2F12D78F|D7E26BC344D7E224|E69AF134B5E29AD5|98BC4D6BF898BC58|
|BD135E2FA6BD1346|12D789A92F12D78F|D7E26BC344D7E224|E69AF134B5E69AD5|98BC4D6BF898BC58|
|BD135E2FA6BD1346|12D789A92F12D78F|D7E26BC344D7E224|E29AF134B5E69AD5|98BC4D7BF898BC58|
```

```
|-----|-----1---|-----|-----|--4-----|
|-----|-----|-----1---|-----8-----|-----|
|-----|-----|-----8-----|-----|-----4-----|
|-----|-----|-----|-----|-----|
```

```
|-----|-----|-----4-----|-----|-----|
|-----2---|-----|-----4-----|-----|-----|
|-----2---|-----4-----|4-----|-----|-----|
|-----|-----|-----|-----|-----|
```

```
|-----|-----|-----1---|-----8---|-----|
|-----|-----|-----8---|-----|-----|
|-----|-----|-----|-----2-----|-----|
|-----1---|-----|-----|--4-----|-----|
```

```
|-----|2-----|48----4--2---|4---12-----|--8--82---1-|
|---98-----|2---2-8---4--|---12-----|4--2--2-8---|--1--8---2---|
|-----4---|2---1-----|---12-----|4--2--2-8---|--4-----|
|---1-4---2---1|-----|---8-----|--2---8---4-|---4-----9-|
|--2---1---4--|-----48-|1-4---2---1---|-----|-----8---|
```

The characteristic leads to near-collisions with a Hamming distance of 5 for Keccak-224, and 8 for Keccak-256. The probability of each one of the first three transitions is  $2^{-12}$ . The probability of the final transition is  $2^{-21}$  for Keccak-224 and  $2^{-23}$  for Keccak-256. The total probability is  $2^{-57}$  for Keccak-224 and  $2^{-59}$  for Keccak-256.

**Characteristic 2:** A 4-round characteristic leading to near-collisions on Keccak-224 and Keccak-256

bits ordered from left to right, where each block is given in hexadecimal using the little-endian format.

```
M1=  
C4F31C32 4C59AE6D 5D19F0F4 25C4E44B D8853032 8D5E12F2 BB6E6EE2 27C33B1E 6C091058 EB9002D5  
3BA4A06F 4A0CC7F1 CCB55E51 8D0DD983 2B0A0843 9B21D3B0 53679075 526DDED2 48294844 6FF4ED2C  
1ACE2C15 471C1DC7 D4098568 F1EBF639 EAF7B257 09FDAE87 688878E6 4875EB30 C9C32D80 3C9E6FCB  
3C2ABCFA E6A4807B 2AB281B8 812332B3
```

```
M2=  
A4D30EF7 80BB8F69 90C048DF EB7213B9 A6650424 3A65F63E 8C268881 B651B81F AADAF3C EE2CA5C3  
DB78EAC2 C8EAE779 442F9C35 3221E287 B3017A5A 90790712 1B1C8BDC E08B10A8 9A9D25CA 1BE7AAAC  
4E2F3E9C 73717DAD 5566015A A198CFB9 5A1CA8C2 A0E3348A AE6C0BB1 3980F9E4 A4FA8B91 6E81A989  
89A9BCAA E12BF1F1 30EF9595 812E8B45
```

```
Output=  
61FB1891 F326B6D5 24DD94DF 73274984 05DA9A1D 3FD359B9 78B8393B F2E7990B
```

The messages were found using the target difference algorithm on the target difference given by Characteristic 1.

**Collision 1:** A collision for 4-round Keccak-256

```
M1=  
FAC7AC69 2710BE04 8462C382 7ABF1BF9 D065CD30 DB64DEB8 1410CD30 C837D79B 22E446B7 31E9BD55  
A6B2074C C86E32CC DE50A10A F7BAAA58 D96CBC88 9FBD75F6 5E0D735A D22AA663 16A574AA 7DB08692  
558AB029 109B4D30 86CE5DCA 13A295C7 E7C9D94B 648794D2 62EE3CF8 69439337 8CAB9F15 AC7C3267  
90F41CBE A20E6893 B4781F24 0BA37647 F29A67A0 81F628D0
```

```
M2=  
CESFBC81 47710FCC 462C92E0 48F5D2CF F92F6EC3 053E64E1 570780B9 F838EC54 8F74809F 66B4AC6F  
70DD1843 BF34F0C5 5010C89A D8791148 D5CC073E 3239AEB3 7DF48D79 0EC7767B FB081604 AFA975B9  
F8EFAE0F ED793473 479E931C F2F80A74 7192D08F 5EB5AB27 F1CAC04E F394232D 48656B2A A3471644  
DB74E60A 05FB3B18 41DC27C3 8384BF53 32534C3E 811C00B5
```

```
Output=  
826B10DC 0670E4E1 5B510CDA AB876AA8 B50557ED 267932FB AA4D38E8
```

The messages were found using the target difference algorithm on the target difference given by Characteristic 1.

**Collision 2:** A collision for 4-round Keccak-224

M1=  
23296F07 44536A2B 16E1E363 09B509F9 639CA324 2B834133 61457E6D 9CF07597 6797B3D4 D1715ABA  
6D8F4F9F 70D12920 E014BB37 54C32ADE 6117B3FB 30114566 4BA7D70A 00F055F0 71CFDD4 B53F2563  
E223A16D CC8DDAC4 7A59836B A53FBDDE 9FFEC45F 6A3476DC 7349BB92 56AF6E92 83866932 56624032  
A936E410 60AC00FA 7E7C61F9 81583CAC

M2=  
49D48DE2 9FA843CA 747C88E0 55425134 098CA5B3 C97DC68A B82BC6FD 0F864996 26B13425 D9F73B75  
932CD02F FB12E036 47706100 9DEFFFE4 79435F9C DA727EFO D9CA67C6 520BE2D1 19CF3933 3136D1A9  
EEBEA9DD 150CA247 D494BF4A 492EFB26 11CB4C8D F5A10A05 69128FF4 B142742F CA59FE32 4FE68436  
068F76AB 041A673E 461575B5 81AA2A54

Output1=  
407D4466 FEA8B231 EC968181 DF902165 23C219FF 54571D70 2800F506 E818644B

Output2=  
407D4466 FEA8B231 EC928181 FF902165 23C019FF 1C571D74 2800F516 E810656B

The messages were found using the target difference algorithm on the target difference given by Characteristic 2.

**Near-Collision 1:** A near collision with Hamming distance of 10 for 5-round Keccak-256

M1=  
7DBC1AA9 62A70B2A C2BDAF81 4A4D484B 672F6FAF ED312C83 24BC1974 16946039 6B46EDF6 1AE571A0  
EDA59D6E 7561766D 8F0B4C57 3C05C569 715B7DF9 53F81761 F6D43507 6E040495 9B5C08AB 5130BA66  
22AF7F5C 755840F2 2E893F59 4C4A730F 8C4F425D 182F8D00 E98515ED E29251AD 853AB863 DC46A7AC  
9FB7BB08 14767EFC 5345C7AF AA774E81 8A01A570 81D65453

M2=  
5659C936 AF3BA787 809C1CE6 B287F81B E0A5E769 ECCEB8A0 72506F44 1A1B2A02 EE9AE408 D16A9358  
BF03C4D6 90845C46 0C0441CC 8203EA8D 6D122EB1 9193F64F 55C3A6A7 47377ED6 D26E806F DEC2CBF8  
A3B8949E A91B248D 420B13BC BEAB4166 EE348CF6 DB6CCD82 122F6BDA 2FBFA7E4 75E8A429 F397BC46  
7E9DE824 6A973A22 371FD02D 92035083 267D1C7A 812EDE70

Output1=  
85373497 97D871C2 FBD0A823 584C0ED4 C1B3BF4F BC408766 0584B08D

Output2=  
85373497 97D871C2 FBD0A823 784C0ED4 E1B1BF5F BC408776 0584B08D

The messages were found using the target difference algorithm on the target difference given by Characteristic 2.

**Near-Collision 2:** A near collision with Hamming distance of 5 for 5-round Keccak-224